Commun. Math. Phys. 93, 437–451 (1984)

Communications in Mathematical Physics
© Springer-Verlag 1984

Instantons in Two and Four Dimensions

M. F. Atiyah

Mathematical Institute, University of Oxford, 24–29 St. Giles, Oxford OX1 3LB, United Kingdom

Abstract. It is shown that Yang-Mills instantons in four dimensions can naturally be identified with the instantons of a two-dimensional theory with values in the loop group.

1. Introduction

Because of the daunting difficulties involved in attempting to quantize realistic physical gauge-theories in four-dimensional space-time considerable attention has been given to certain two-dimensional (2D) models, which it is hoped share some of the important qualitative features of the four-dimensional (4D) theories. In particular pure Yang-Mills theory in 4D is compared with the CP_n-models in 2D. Both theories are conformally invariant and possess instantons, and this provides a basis for obvious analogies.

The purpose of this paper is to strengthen the analogy concerning the instantons in these two theories. Essentially we shall show (at least for G a classical group and probably for all G) that Yang-Mills instantons in 4D can be naturally identified with (i.e. have the same parameter space as) the instantons in 2D for the theory in which the complex projective n-space CP_n is replaced by the *infinite-dimensional* manifold ΩG of loops on the structure group G. Such a theory is not as bizarre as it appears because ΩG is well-known to share most of the important properties of CP_n and it arises naturally in many contexts.

A natural identification between the instantons of two different theories suggests that there might be a close relation between the two field-theories involved. Our result therefore indicates that it would be worth exploring the two-dimensional theory for ΩG-valued fields, and that this might provide a bridge between the CP_n-models in 2D and Yang-Mills theory in 4D.

It is well-known that the CP_n-instantons are given by holomorphic (or rational) maps $CP_1 \to CP_n$, where $CP_1 = R^2 \cup \infty$ is the conformal compactification of R^2. This depends on the fact that CP_n is a Kähler manifold. Now ΩG is an infinite-dimensional Kähler manifold [15, 17] and so ΩG-instantons are also given by holomorphic maps

$$f: CP_1 \to \Omega G. \tag{1.1}$$

On the other hand Yang-Mills instantons on R^4 can be constructed by twistor methods and are described by suitable holomorphic data. It turns out that this holomorphic data amounts essentially to a holomorphic map f as in (1.1). Moreover a recent result of Donaldson [7], refining earlier work on Yang-Mills instantons, implies that all maps f arise in this way. Actually Donaldson gives the proof only for the classical groups but it seems likely that his result holds for all G.

In Sect. 2 we review the basic properties of ΩG and show that holomorphic maps f as in (1.1) correspond essentially to holomorphic bundles over $CP_1 \times CP_1$ with group G^c (the complexification of G). In Sect. 3 we recall how Yang-Mills instantons on R^4 can be re-interpreted, by twistor methods, in terms of holomorphic bundles on CP_3 with a "real structure" or more simply, following Donaldson, as holomorphic bundles on CP_2. Since CP_2 is closely related to $CP_1 \times CP_1$ (both being compactifications of C^2) this leads easily to our main result (Theorem 1) identifying ΩG-instantons in 2D with Yang-Mills instantons in 4D.

In Sect. 4 we specialize Theorem 1 to axially symmetric instantons (with "axis" $R^2 \subset R^4$) and find (Theorem 2) that these correspond to holomorphic maps of CP_1 into one of the (finite-dimensional) Kähler manifolds which occur as homogeneous spaces of G. For example when $G = \mathrm{SU}(2)$ we get just holomorphic maps $CP_1 \to CP_1$. These results are of interest in connection with the study of magnetic monopoles and in Sect. 5 we discuss this aspect, and in particular the relation with another recent result of Donaldson [8].

2. The Holomorphic Structure of ΩG

We shall begin by reviewing briefly the basic facts about ΩG. For fuller details we refer to [17] – see also [15, 16].

For any compact Lie group G the loop space ΩG consists of all based maps of the unit circle S^1 ($|z| = 1$) into G, $f: S^1 \to G$ with $f(1) = 1$. To be precise one should of course specify the class of maps f to be used. For many purposes it is immaterial which class is used provided sufficient differentiability is assumed. If f is taken in the Sobolev space H^1 (having first derivatives square integrable) then ΩG becomes a Hilbert manifold, which is sometimes convenient. For our purposes it will be sufficient to take f to be C^∞ or even real analytic.

The simplest case is of course when $G = U(1)$. Then ΩG has components indexed by the winding number and each component can (by taking logarithms) be identified with the space of real-valued functions $f: S^1 \to R$, $f(1) = 0$. The Fourier series expansion of such a function is

$$\phi = \sum_{-\infty}^{\infty} a_n z^n, \qquad a_{-n} = \bar{a}_n, \qquad \sum a_n = 0 \tag{2.1}$$

and so is entirely determined by the Fourier coefficients $\{a_n\}$ for $n > 0$. In this way each component of ΩG becomes an infinite-dimensional complex vector space.

For non-abelian G the situation is more complicated because ΩG is no longer linear. However it is an infinite-dimensional manifold and infinitesimally one can again use a Fourier series decomposition to introduce complex coordinates. More precisely the tangent space to ΩG at the base point [the constant loop $f(z)=1$] consists of Lie-algebra-valued functions on S^1, and can be represented by a Fourier series

$$\phi=\sum_{-\infty}^{\infty} a_n z^n\,, \qquad a_{-n}=-a_n^*\,, \qquad \sum a_n=0\,, \tag{2.2}$$

where a_n lies in the complexification of the Lie algebra of G. Thus, when $G=U(m)$, the coefficients a_n are *complex* $m\times m$ matrices and a_n^* is the transposed conjugate matrix. Thus the tangent space to ΩG, at the base point, becomes a complex linear space. Now using the obvious group structure on ΩG we can transport this complex structure to all tangent spaces. It turns out that this infinitesimal (or "almost") complex structure is actually integrable so that ΩG is an infinite-dimensional complex manifold. An alternative and more useful description of the complex structure will be explained shortly.

To define a hermitian metric on ΩG it is again enough to specify it at one point and then to translate it by the group action. There are several natural metrics one can introduce but the most natural is, in some ways, the one given relative to the Fourier coefficients of (2.2) by

$$\sum_{n>0} n \text{ trace } (a_n a_n^*)\,, \tag{2.3}$$

where a_n is here viewed as a matrix (or more intrinsically one views "trace" as the Killing form). The reason why (2.3) is the natural metric lies in the fact that it is a *Kähler* metric. The associated symplectic form is given by

$$(\phi,\psi)=\frac{1}{2\pi}\int_0^{2\pi}\langle\phi',\psi\rangle d\theta\,, \tag{2.4}$$

where $\langle\,,\rangle$ is given by the Killing form and $\phi'=\dfrac{d\phi}{d\theta}$.

If G is a simple and simply-connected Lie group then its first few homotopy groups are:

$$\pi_1(G)=\pi_2(G)=0\,, \qquad \pi_3(G)\cong Z \text{ (the integers)}.$$

Since $\pi_i(G)\cong\pi_{i-1}(\Omega G)$ it follows that

$$\begin{cases}\Omega G \text{ is connected, simply connected and}\\ \pi_2(\Omega G)\cong Z\,.\end{cases} \tag{2.5}$$

In particular it follows that the second homology $H_2(\Omega G)$ is also isomorphic to the integers. Moreover one can check that the Kähler 2-form integrated over a generating 2-cycle is non-zero (and can be normalized to be 1).

We see therefore that ΩG does indeed share most of the important properties of CP_n, except of course finite-dimensionality.

To understand this analogy further it is helpful to recall that CP_n can be described either as a homogeneous space of $U(n)$ or of $GL(n,C)$:

$$CP_n=U(n)/U(1)\times U(n-1)=GL(n,C)/H\,,$$

where H consists of all matrices of the form h_{ij} with $h_{i1}=0$ for $i>1$. The first description shows that it is compact while the second exhibits the complex structure. The unitary description occurs naturally when we consider the orbit of the matrix

$$\begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & & \\ & & \alpha_3 & \\ & & & \alpha_n \end{bmatrix} \quad \alpha_i \text{ real}, \qquad \alpha_2 = \alpha_3 = \ldots = \alpha_n \neq \alpha_1 \tag{2.6}$$

under conjugation by $U(n)$. More generally, for any compact Lie group G, the orbit of any vector α in the Lie algebra of G under the adjoint representation is a homogeneous space M of the form

$$M = G/G(\alpha), \tag{2.7}$$

where $G(\alpha)$ is the centralizer of α (elements of G commuting with α).

This also has a complex description as

$$M = G^c/P, \tag{2.8}$$

where P is a suitable complex subgroup of the complexification G^c of G. These spaces all have natural Kähler metrics and the associated symplectic 2-forms are given by a general procedure due to Kirillov [12]. Note however that the second homology group $H_2(M)$, unlike the case of CP_n, may have several generators, e.g. for a general diagonal matrix (2.6) we get $n-1$ generators.

Let us now return to our loop space ΩG and explain how it also has two different expressions. For this we shall introduce the group $\mathscr{G}$ of *free loops* on G, i.e. all maps $f: S^1 \to G$ without restriction on base points. The constant loops give G as a subgroup of $\mathscr{G}$ and clearly the coset space can be identified with ΩG:

$$\Omega G = \mathscr{G}/G. \tag{2.9}$$

This is to be interpreted as the analogue of (2.7). The best justification for this analogy is to consider $\mathscr{G}$ as the group of gauge transformations for a G-bundle over S^1. Then $\mathscr{G}$ naturally acts on the space $\mathscr{A}$ of connections (gauge potentials) and ΩG appears as the orbit of the trivial connection. Moreover this affine action of $\mathscr{G}$ on $\mathscr{A}$ can be interpreted in terms of the co-adjoint action of a central extension of $\mathscr{G}$ [18]. By the Kirillov procedure ΩG thus acquires a natural symplectic form and this can be identified with (2.4).

The complexification $\mathscr{G}^c$ of $\mathscr{G}$ consists of maps $f: S^1 \to G^c$, and this has a subgroup $\mathscr{P}$ consisting of maps f which extend to *holomorphic* maps of the closed unit disc $|z| \leq 1$ to G. The analogue of (2.8) is then true, namely

$$\Omega G = \mathscr{G}^c/\mathscr{P}, \tag{2.10}$$

and this endows ΩG with the complex structure defined infinitesimally from (2.3). The identification (2.10) amounts to the two group-theoretical statements:

(i) $$\mathscr{G}\mathscr{P} = \mathscr{G}^c,$$

(ii) $$\mathscr{G} \cap \mathscr{P} = G.$$

Essentially (i) asserts the existence of a solution for a certain non-linear boundary-value problem while (ii) gives its uniqueness modulo constants.

Using (2.10) it follows that, for any (finite-dimensional) complex manifold X, a holomorphic map $f: X \to \Omega G$ will, relative to some open covering $\{U_i\}$ of X, be given by holomorphic maps $f_i: U_i \to \mathcal{G}^c$, which agree modulo $\mathcal{P}$, i.e. for each pair i, j, $f_i f_j^{-1}: U_i \cap U_j \to \mathcal{P}$. Moreover, to say that f_i is holomorphic, simply means that the corresponding map, $F_i: U_i \times S^1 \to G^c$, given by $F_i(x, z) = f_i(x)(z)$, is holomorphic (in some neighbourhood in $U_i \times C$).

Now holomorphic maps $f: S^1 \to G^c$ naturally define holomorphic G^c-bundles over CP_1. Explicitly we decompose CP_1 as the union of the two discs D_0 ($|z| \leqq 1$) and D_∞ ($|z| \geqq 1$). Then we take trivial G^c-bundles over neighbourhoods of D_0, D_∞ and identify them over the intersection using f. If f also depends holomorphically on a parameter space X then we will get a holomorphic bundle over $X \times CP_1$. We proceed to put this into more precise form. It will be convenient to introduce base-points $x_0 \in X$, $\infty \in CP_1$, $x_0 \times \infty \in X \times CP_1$ and $1 \in \Omega G$ [here 1 is the constant loop $f(z) = 1$]. Finally a *based* bundle will mean a bundle together with a trivialization of the fibre over the base point (i.e. an identification of this fibre with the structure group). The precise result we are after is then the following:

Proposition (2.11). *Let X be a compact connected complex manifold with base point. Then there is a natural equivalence between*

(i) *based holomorphic maps $X \to \Omega G$,*

(ii) *based holomorphic G^c-bundles on $X \times CP_1$ which are trivial on $x_0 \times CP_1$ and $X \times D_\infty$.*

Proof. Given a holomorphic bundle as in (ii) we can pick an open covering $\{U_i\}$ of X so that the bundle is trivial over each $U_i \times D_0$ (and also by hypothesis on $X \times D_\infty$).

Choices of trivialization over $X \times D_\infty$ correspond to holomorphic maps $X \times D_\infty \to G^c$, and since X is compact and connected these have to be constant on the X-factor and depend only on D_∞. Moreover the trivialization over the base-point $x_0 \times \infty$ extends to a unique trivialization over $x_0 \times CP_1$ and this restricts to a unique trivialization on $x_0 \times D_\infty$. Thus we have a definite choice of trivialization on all of $X \times D_\infty$. On each $U_i \times D_0$ we pick any trivialization consistent with that on $x_0 \times CP_1$ (if $x_0 \in U_i$). The holomorphic bundle is then given by based holomorphic maps, $f_i: U_i \to \mathcal{G}^c$, which agree modulo $\mathcal{P}$, and so define a based holomorphic map, $f: X \to \mathcal{G}^c/\mathcal{P} = \Omega G$. It is easy to see that this is independent of the choice of open covering $\{U_i\}$ and establishes a natural one-one correspondence between (i) and (ii).

We are interested in the special case of (2.11) when $X = CP_1$. Note that in this case the topological classification of either (i) or (ii) (for a simple group G) is by a single integer. In (i) this is the "degree" of the map f, defined as the multiple of the generator of $H_2(\Omega G)$ represented by f or equivalently by the integral over CP_1 of $f^*\omega$ where ω is the (normalized) Kähler form on ΩG. In (ii) the integer appears as the 4-dimensional characteristic class of the bundle on $CP_1 \times CP_1$. For example if $G = U(n)$ this is the second Chern class.

Continuing with the case $X = CP_1$, we want next to weaken the assumption in (2.11) (ii) that the bundle is trivial on $X \times D_\infty$, replacing it by assuming triviality only on $X \times \infty$. Now, for bundles on CP_1, triviality is an open condition and so any bundle on $CP_1 \times D_\infty$ which is trivial on $CP_1 \times \infty$ is automatically trivial on $CP_1 \times D'$ for some disc D' with $\infty \in D' \subset D$. This follows easily from the "Birkhoff stratification" of ΩG described in [16, 17], since the unique open stratum corresponds to trivial bundles.

Since a rescaling will take D' into D_∞ we can get back to the situation of (2.11). To formalize this we introduce the following notation:

$\mathscr{H}_k(CP_1, \Omega G)$ = based holomorphic maps $CP_1 \to \Omega G$ of degree k,

$\mathscr{M}_k(CP_1 \times CP_1, CP_1 \vee CP_1; G^0)$ = based isomorphism classes of holomorphic G^c-bundles over $CP_1 \times CP_1$, trivial over $CP_1 \vee CP_1$ (the union of the axes) and with characteristic class k.

Both $\mathscr{H}_k$ and $\mathscr{M}_k$ are naturally complex manifolds. In fact general results of algebraic geometry [13] imply that $\mathscr{M}_k$ is actually a (non-compact) algebraic variety (see Sect. 3) and Proposition (2.11) identifies $\mathscr{H}_k$ with an open set of $\mathscr{M}_k$. Moreover the complex structure on $\mathscr{H}_k$ induced by this identification is its natural one as a space of holomorphic maps, $CP_1 \times S^1 \to G^c$. Now, as pointed out above, any holomorphic bundle represented by a point ξ of $\mathscr{M}_k$ is trivial on $CP_1 \times D'$ for some disc D' given by $|z| \geqq \varrho$. The infimum of such ϱ is in fact continuous in ξ and so defines a continuous non-negative function, $\phi: \mathscr{M}_k \to R$. Let us put $\mathscr{M}_k^\mu = \phi^{-1}[0, \mu)$, then $\mathscr{M}_k^\mu$ is an open set in $\mathscr{M}_k$ and (2.11) gives a homeomorphism

$$\mathscr{M}_k^1 \cong \mathscr{H}_k. \tag{2.12}$$

On the other hand the rescaling $z \to \mu z$ (with μ real and positive) induces an action of the multiplicative group on $\mathscr{M}_k$ which we also denote by $\xi \to \mu\xi$. From its definition we see that ϕ is equivariant for this action, i.e. $\phi(\mu\xi) = \mu\phi(\xi)$. This implies that μ induces a homeomorphism, $\mathscr{M}_k^1 \cong \mathscr{M}_k^\mu$. We can now "stretch" $\mathscr{M}_k^1$ into $\mathscr{M}_k$ by using the map $\xi \to (1-\phi(\xi))^{-1}\xi$, whose inverse is $\eta \to (1+\phi(\eta))^{-1}\eta$. Together with (2.12) this establishes the homeomorphism

$$\mathscr{M}_k \cong \mathscr{H}_k. \tag{2.13}$$

Remark. The equivalence (2.12) is naturally complex analytic but the stretching which yields (2.13) is not. We could make (2.13) a diffeomorphism by smoothing the continuous function ϕ but the holomorphic structures on $\mathscr{M}_k$ and $\mathscr{H}_k$ are essentially different, in the same way as the complex plane differs from the unit disc.

3. Yang-Mills Instantons

In this section we shall review the known results about Yang-Mills instantons, including the important recent result of Donaldson [7]. Combined with the results of Sect. 2 this will then lead to our main conclusion, Theorem 1.

A Yang-Mills instanton over R^4, with group G, is a G-connection (or potential) whose associated curvature F (or field) satisfies

$$\begin{cases} *F = -F \\ \int |F|^2 < \infty . \end{cases} \tag{3.1}$$

Here we have chosen the minus sign (giving anti-self dual solutions) to fit the usual orientation conventions of complex structure. The normalized action

$$\frac{1}{8\pi^2}\int |F|^2 \tag{3.1}$$

is then a non-negative integer k, called the instanton number.

If we introduce complex coordinates (z_1, z_2) so that we identify R^4 with C^2 then F can be decomposed into types:

$$F = F^{2,0} + F^{1,1} + F^{2,0}, \qquad F^{2,0} = -(F^{0,2})^* .$$

The anti-self-dual condition $*F = -F$ then breaks up into two parts:

$$\begin{cases} \text{(a)} \ F^{0,2} = 0, \\ \text{(b)} \ F^{1,1} \wedge \omega = 0, \end{cases} \tag{3.2}$$

where $\omega = \sum_{j=1}^{2} dz_j \wedge d\bar{z}_j$ is the standard 2-form. Condition (a) is just the integrability condition for a holomorphic structure on the G^c-bundle, while (b) amounts to a supplementary unitary condition involving essentially the choice of a suitable metric in the fibres.

Locally there are many independent solutions of (3.2) but globally, under suitable conditions, there is a unique solution of (b) for each solution of (a). This has been proved by Donaldson [6], when C^2 is replaced by a *compact* Kähler surface, by direct analytical methods. For the case of C^2, in which we are now interested, Donaldson [7] has given a more algebraic argument which relies on the explicit ADHM construction [1, 3], and applies to any classical group G. To describe Donaldson's result precisely let us first define the following parameter spaces:

$M_k(R^4, G)$ = space of k-instantons over R^4, with group G, modulo based gauge equivalence,

$\mathcal{M}_k(CP_2, CP_1; G^c)$ = space of based isomorphism classes of holomorphic G^c-bundle on CP_2 which are trivial on CP_1.

Based gauge equivalence means that we only allow gauge transformations which $\to 1$ at ∞. Equivalently if we compactify R^4 conformally to S^4, then we use $\infty \in S^4$ as base point. Similarly CP_2 is regarded as a compactification of C^2, with CP_1 as the "line at ∞" and we take any point on CP_1 as base point. Donaldson's result is then:

Proposition (3.3) [Donaldson]. *For any classical group G, there is a natural diffeomorphism*

$$M_k(R^4, G) \cong \mathcal{M}_k(CP_2, CP_1; G^c).$$

Remarks. (1) What is noteworthy about this result is that the space $\mathcal{M}_k$ is a purely holomorphic (even algebraic) object not involving any real or unitary structure.

(2) The usual twistor approach to instantons, as explained in [1], uses *all* the complex identifications $R^4 = C^2$ and ends up with a description of instantons in terms of holomorphic bundles on CP_3, with *extra reality constraints*.

(3) The map $M_k \to \mathcal{M}_k$ is easy to define explicitly, and it is not hard to see that it is *injective*. The force of Donaldson's result is that it is *surjective*. On the other hand

this is purely an existence theorem and the inverse map $\mathcal{M}_k \to M_k$ is *not* constructed explicitly. This means there is no elementary procedure to assign to each point of the parameter space $\mathcal{M}_k$ an explicit instanton.

(4) M_k is naturally just a differentiable manifold, but $\mathcal{M}_k$ is a complex manifold. Thus (3.3) endows M_k with a complex structure. In fact, since this depended on a choice of identification $R^4 = C^2$, we have a whole 2-parameter family (parameterized by S^2) of such complex structures on $\mathcal{M}_k$.

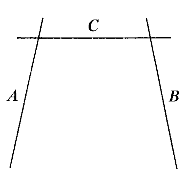
(5) If we use all gauge transformations on S^4, instead of based transformations, we get a smaller parameter space which is M_k/G. For example if $G = \mathrm{SU}(2)$, $\dim M_k = 8k$, while $\dim M_k/G = 8k - 3$. Moreover, for general G, M_k is a manifold but M_k/G has singularities arising from reducible instantons.

(6) In algebraic geometry good parameter (or moduli) spaces of algebraic bundles exist only if one restricts to *stable* bundles. However, over CP_2 the condition that bundles are trivialized over CP_1 essentially ensures stability.

Comparing (3.3) with the results of Sect. 2 we see that here we have been considering holomorphic bundles over CP_2 (trivial on CP_1), while in Sect. 2 we considered holomorphic bundles on $CP_1 \times CP_1$ (trivial on $CP_1 \vee CP_1$). Now CP_2 and $CP_1 \times CP_1$ are birationally equivalent (given by the usual stereographic projection of the quadric surface $CP_1 \times CP_1$ onto CP_2). More precisely there is a diagram

$$\begin{array}{ccc} & Y & \\ {}^{\alpha}\swarrow & & \searrow^{\beta} \\ CP_2 & & CP_1 \times CP_1 \end{array} \qquad (3.4)$$

where Y is a third algebraic surface. The map α collapses two disjoint copies of CP_1 in Y to distinct points A, B on $CP_1 \subset CP_2$. The map β collapses a third CP_1 in Y to a point C. On Y these three copies of CP_1 form a configuration as indicated



(3.5)

the lines being indicated by the points into which they collapse under α and β. In such a situation quite general algebraic geometric arguments ensure that a holomorphic bundle on Y, with a trivialization on the configuration of lines (3.5), can be pushed down by α to give a bundle on CP_2 trivialized at A and B. Equally it can be pushed down by β to give a bundle on $CP_1 \times CP_1$ trivialized at C. Set-theoretically the pushing down process is clear, the only non-trivial point is that one gets a locally trivial holomorphic bundle and this is equivalent to constructing holomorphic trivializations on Y in the neighbourhood of the curve to be collapsed. The proof that this is possible is in two stages. First one constructs, step by step, a formal Taylor series (in the normal coordinates), and then one appeals to a general theorem of Grothendieck (Éléments de Géométrie Algebrique III, Publ. Math. Inst. des Hautes Etudes Sci. No. 11) asserting that such Taylor series are generated by algebraic functions.

Thus the birational correspondence (3.4) leads to a natural identification of parameter spaces:

$$\mathcal{M}_k(CP_2, CP_1; G^c) \cong \mathcal{M}_k(CP_1 \times CP_1, CP_1 \vee CP_1; G^c). \tag{3.6}$$

Remark. More generally the same argument shows that we can replace CP_2 by any algebraic compactification of C^2. Thus the space $\mathcal{M}_k$ in (3.6) can properly be thought of as the parameter space of algebraic G^c bundles over C^2 with a trivialization at ∞.

Combining (2.13), (3.3), and (3.6) we end up finally with our main result:

Theorem 1. *For any classical group G and positive integer k, the following two spaces are diffeomorphic:*

(1) *the parameter space of Yang-Mills k-instantons over R^4 with group G, modulo based gauge transformations,*

(2) *the parameter space of all based holomorphic maps*

$$CP_1 \to \Omega G$$

of degree k.

One possible application of Theorem 1 would be to prove the conjecture made in [4] concerning the topology of the instanton parameter spaces M_k as $k \to \infty$. More precisely it was conjectured in [4] that the natural inclusions $M_k \to \Omega^3 G$ induced isomorphisms in all homotopy groups up to dimension q provided $k > k_0(q)$. Here Ω^3 stands for the space of based maps $S^3 \to G$. On the other hand it is a theorem of G. B. Segal [19] that the space of based holomorphic maps $CP_1 \to CP_n$ of degree k approximates the space $\Omega^2(CP_n)$, in the same sense, as $k \to \infty$. Moreover, this result of Segal's can be generalized to other homogeneous spaces besides CP_n, and it seems likely that the same methods will extend to the case of ΩG. Together with Theorem 1 this would then prove the conjecture of [4], at least for the classical groups.

It is generally believed, though no-one has yet proved, that for $G = \mathrm{SU}(2)$ the full Yang-Mills equations on S^4 have no solutions except instantons (and anti-instantons). Theorem 1 then suggests that one might expect the same to be true for harmonic maps

$$CP_1 \to \Omega(\mathrm{SU}(2)). \tag{3.7}$$

The corresponding result for maps $CP_1 \to CP_1$ is easy and well-known, but for (3.7) the problem looks interestingly non-trivial.

4. The Axially-Symmetric Case

In this section we shall specialize Theorem 1 to the case of axially symmetric instantons. This leads to an especially simple answer (Theorem 2) and, as we shall explain in Sect. 5, it has an interpretation in terms of magnetic monopoles.

We begin with a few further remarks on the geometry of the loop group ΩG. In the first place there is an action of the circle group S^1 on ΩG obtained by rotating loops. Since ΩG consists of *based* loops this action is defined by

$$(\lambda f)(\mu) = f(\lambda\mu) f(\lambda)^{-1}, \quad \text{for} \quad \lambda \in S^1 \quad \text{and} \quad f \in \Omega G.$$

In particular f is a *fixed point* of this action if and only if $f(\mu)=f(\lambda\mu)f(\lambda)^{-1}$ or $f(\mu)f(\lambda)=f(\mu\lambda)$, i.e. f is a *homomorphism* $S^1 \to G$. If α is any such homomorphism then so are all its conjugates, and these are naturally parametrized by the homogenous space $\Gamma_\alpha = G/G(\alpha)$, where $G(\alpha)$ is the centralizer of the one-parameter group α. Thus the fixed point set Γ of the action of S^1 on ΩG is the disjoint union $\Gamma = \bigcup_{\alpha \in \Lambda} \Gamma_\alpha$, where α runs over the set Λ of conjugacy classes of homomorphisms of S^1 into G. This indexing set Λ can naturally be identified with the integer lattice in the Lie algebra of G (kernel of the exponential map) modulo the action of the Weyl group. For example when $G = \mathrm{SU}(2)$, the set Λ becomes the non-negative integers and $\Gamma_0 = \text{point}$, $\Gamma_n = CP_1$ for $n \geqq 1$.

As we have pointed out before the homogeneous spaces $G/G(\alpha)$ are all Kähler manifolds. In fact they appear naturally as the Kähler submanifolds Γ_α of ΩG. Note that, simply as a complex manifold, each of the homogeneous spaces appears many times in ΩG. The different α corresponds to different choices of the Kähler metric. For example when $G = \mathrm{SU}(2)$ the Kähler metric on Γ_n is just n times the Kähler metric on Γ_1. However for larger groups the lattice has several generators and the Kähler metrics are not all proportional.

The group G acts naturally by conjugation on ΩG, the loop $f(\lambda)$ getting conjugated by $g \in G$ into $gf(\lambda)g^{-1}$. In particular for any homomorphism $\alpha: S^1 \to G$, we get an induced action of S^1 on ΩG, the loop $f(\lambda)$ being transformed by $\mu \in S^1$ into the loop $\alpha(\mu)f(\lambda)\alpha(\mu)^{-1}$. Consider now the set $\Gamma^\alpha \subset \Omega G$ of loops for which this action of S^1 coincides with the *inverse* of the other action of S^1 given by loop rotation, so that

$$f \in \Gamma^\alpha \Leftrightarrow \alpha(\mu)^{-1} f(\lambda)\alpha(\mu) = f(\mu\lambda) f(\mu)^{-1}. \tag{4.1}$$

Putting $h(\mu) = \alpha(\mu)f(\mu)$ and recalling that α is a homomorphism, so that $\alpha(\lambda\mu) = \alpha(\lambda)\alpha(\mu)$, (4.1) becomes

$$f \in \Gamma^\alpha \Leftrightarrow h(\lambda)h(\mu) = h(\lambda\mu). \tag{4.2}$$

Thus $h: S^1 \to G$ is also a homomorphism, i.e. $h \in \Gamma$, and so (4.2) implies that

$$\Gamma^\alpha = \alpha^{-1}\Gamma. \tag{4.3}$$

In particular the component $(\Gamma^\alpha)_1$ of Γ^α containing the base point $1 \in \Omega G$ is just a left translate of the homogeneous space Γ_α:

$$(\Gamma^\alpha)_1 = \alpha^{-1}\Gamma_\alpha. \tag{4.4}$$

After this digression we return to the consideration of Yang-Mills instantons on R^4. We fix an orthogonal decomposition

$$R^4 = R^2 \oplus R^2, \tag{4.5}$$

and consider rotation about the origin in the second R^2, extended trivially to the first factor. Thus the first R^2 becomes the "axis" of the rotation.

An S^1-invariant connection over R^4, relative to this action of S^1 on R^4, will mean a principal G-bundle P with connection A *together with a lifting of the action of* S^1 to an action on P preserving A. Isomorphisms of two such S^1-invariant connections will always be required to commute with the S^1-actions on the bundles. Since two different liftings of the S^1-action differ by bundle automorphisms preserving A a lifting is determined by its value over any base point on the axis (including ∞). At such a base point the S^1-action determines a homomorphism $\alpha: S^1 \to G$. If we work with based bundles then α is an invariant, otherwise only its conjugacy class is an invariant. For brevity we shall refer to α as the *type* of the invariant instanton.

We now want to apply Theorem 1 to S^1-invariant instantons. The only points that require special note are the following.

1. We choose our complex coordinates compatible with the decomposition (4.5)

2. A holomorphic bundle on $CP_1 \times CP_1$ which is trivial on $CP_1 \times \infty$ and which is invariant under the S^1-action on the second factor is necessarily trivial on the whole of $CP_1 \times (CP_1 - 0)$. This follows because S^1-invariance complexifies to give C^*-invariance, and triviality near $CP_1 \times \infty$ then transports by C^* to give triviality on $CP_1 \times (CP_1 - 0)$. This means we can work directly with Proposition (2.11), and that we do not need to use the "stretching" employed for Theorem 1.

3. Since the S^1-action α does not preserve the trivialization over the base point, S^1-invariant instantons correspond not to maps of CP_1 into Γ but into Γ^α. Note that we used the inverse of loop rotation in the definition of Γ^α because our base point was ∞ rather than 0. In view of (4.3), translating the base point of ΩG from 1 to α replaces (the 1-component of) Γ^α by Γ_α.

Bearing these points in mind the arguments of Sect. 3 lead to the identification of the parameter space of S^1-invariant instantons:

Theorem 2. *For any classical group G and any homomorphism $\alpha: S^1 \to G$ the parameter space of based S^1-invariant k-instantons of type α is naturally isomorphic to the parameter space of based holomorphic maps $f: CP_1 \to G/G(\alpha)$ of degree k.*

Remark. We have deduced Theorem 2 from Theorem 1. It is also possible [2] to give a direct proof of Theorem 2 without using ΩG, but still using Donaldson's result (3.3).

The degree of a map $f: CP_1 \to G/G(\alpha)$ is defined relative to the Kähler form on $G/G(\alpha)$, which depends on α as pointed out earlier. For example, when $G = \mathrm{SU}(2)$ and α is the homomorphism

$$\alpha(z) = \begin{pmatrix} z^n & 0 \\ 0 & z^{-n} \end{pmatrix},$$

the degree of $f: CP_1 \to G/G(\alpha) = CP_1$ is $2n$ times the usual degree (say p). This means k must be divisible by $2n$

$$k = 2np. \tag{4.6}$$

In general the second homology group of $G/G(\alpha)$ has more than one generator, so that the homology class of f is not determined by k. This means that the parameter spaces in Theorem 2, for given α and k, have several components. These can be characterized by restricting the instanton to the axis of symmetry of R^4. Using the complex description in Sect. 3 this gives a holomorphic bundle ξ on CP_1 (the compactification of the R^2-axis). Since the S^1-action given by α preserves the holomorphic structure, the bundle has its structure group reduced from G^c to

$G(\alpha)^c$. For example, if α is a generic homomorphism, $G(\alpha)$ is a maximal torus T of G and so ξ is a sum of line-bundles. As such it has l integer invariants, where $l = \dim T = \operatorname{rank} G$, which more invariantly can be regarded as an element $\lambda \in H_1(T) \cong H_2(G/T)$. This corresponds to the homology class of the map $f: CP_1 \to G/T$ associated to the instanton in Theorem 2.

The homomorphism α [still in this generic case when $G(\alpha) = T$] maps S^1 into T and so can also be viewed as an element of $H_1(T)$. The degree k of the map f is then given by

$$k = \langle \lambda, \alpha \rangle, \tag{4.7}$$

where $\langle , \rangle$ is the inner product on $H_1(T)$ induced by the Killing form on G (suitably normalized). Formula (4.7) is the generalization of the SU(2)-case (4.6). Moreover essentially similar results hold even when α is not generic, the only difference being that T must be replaced by the connected component of the centre of $G(\alpha)$.

5. Magnetic Monopoles

It is well-known (see for example [10, 11]) that solutions of the self-dual Yang-Mills equations in R^4 which are *independent of* x_4 reduce to solutions of the Bogomolny equations in R^3:

$$\nabla \phi = {}^*F. \tag{5.1}$$

Here the Higgs field ϕ lies in the adjoint representation, $\nabla \phi$ denotes its covariant derivative (with respect to a connection A), F is the curvature of A and $*$ is the duality operator in R^3. Solutions of (5.1) satisfying suitable boundary conditions at ∞ are called (Prasad-Sommerfield) *magnetic monopoles*. As $x \to \infty$ it is assumed that ϕ tends to lie in a fixed G-orbit in the Lie algebra. If α is a point on this orbit then, for x on a sphere S^2 of large radius, the homology class λ of the asymptotic map $\phi_\infty: S^2 \to G/G(\alpha)$ is well-defined and called the *magnetic charge*. For $G = SU(2)$ this is given by a single integer, but in general it is specified by several integers. For fixed α and λ one can then study the parameter space of all magnetic monopoles of *type* α and *charge* λ. Much work has been done in this direction, especially for the case of SU(2) [10, 14].

We can carry out a similar re-interpretation when we replace *translational* invariance by *rotational* invariance. First however we have to replace the Euclidean decomposition

$$R^4 = R^1 \times R^3 \tag{5.2}$$

by the *conformal equivalence*

$$R^4 - R^2 \sim S^1 \times H^3, \tag{5.3}$$

where H^3 is the *hyperbolic* 3-space (of constant curvature -1). If in the decomposition $R^4 = R^2 \oplus R^2$, we use (x, y) coordinates in the first plane and polar (r, θ) coordinates in the second plane, the Euclidean metric of R^4 takes the form

$$ds^2 = dx^2 + dy^2 + dr^2 + r^2 d\theta^2 = r^2 \left\{ d\theta^2 + \frac{dx^2 + dy^2 + dr^2}{r^2} \right\}. \tag{5.4}$$

Since the hyperbolic 3-space H^3 can be represented as the upper half space $z>0$ in (x, y, z) coordinates with metric

$$\frac{dx^2+dy^2+dz^2}{z^2},$$

(5.4) establishes the conformal equivalence (5.3) with the conformal factor being r^2.

Since the self-dual Yang-Mills equations are conformally invariant it follows from (5.3) that solutions of these equations on R^4-R^2 which are *independent* of θ reduce to solutions of the Bogomolny equations (5.1) on the hyperbolic 3-space H^3. Note that the metric of the base manifold enters (5.1) through the *-operator.

Since S^1 is compact (unlike R^1) any finite energy solution of (5.1) on H^3 corresponds to a finite action solution of the self-duality equations on R^4-R^2. In particular therefore the S^1-invariant instantons on the whole of R^4 (or S^4) which we studied in Sect. 4 can be re-interpreted as magnetic monopoles on H^3. Moreover we have chosen our notation with this in mind so that the type α and homological invariant λ of an S^1-invariant instanton do in fact correspond to the type α (limit of Higgs field) and magnetic charge λ of a magnetic monopole on H^3 (these are defined as in the Euclidean case).

The monopoles on H^3 which arise from S^1-invariant instantons on R^4 have the special property that α is *integral*. This is because the solution extends to the R^2-axis in R^4. Examples of solutions with non-integral α, having singularities on the R^2-axis, have been constructed previously [for SU(2)] in [9].

It seems highly likely that all magnetic monopoles on H^3, with integral type α and appropriate asymptotic behaviour, arise from S^1-invariant instantons. If this is so then Theorem 2 implies that the corresponding parameter space for monopoles on H^3 can be identified with the relevant space of holomorphic maps $f: CP_1 \to G/G(\alpha)$. In particular for $G=\mathrm{SU}(2)$ we get just rational maps $f: CP_1 \to CP_1$, i.e. rational functions of one complex variable.

Since $G(\alpha)$ is unchanged if we replace α by any integer multiple, Theorem 2 would also imply that the parameter space of monopoles on H^3 of type $p\alpha$(p an integer) and charge λ is the same as the parameter space for type α and charge λ. Now in Euclidean space one can always rescale the Higgs field by using a dilatation of R^3. In hyperbolic space this is not possible: the curvature gives an absolute scale. However a monopole of type $p\alpha$ on H^3 can be reinterpreted as a monopole of type α on $H^3(p^{-1})$, where $H^3(c)$ denotes the hyperbolic space of constant curvature $-c$ (i.e. H^3 with metric rescaled by c). This follows by returning to R^4 and noting that (5.3) can be replaced by a conformal equivalence

$$R^4-R^2 \sim S^1(c) \times H^3(c), \tag{5.5}$$

where $S^1(c)$ is now the circle of radius c^{-1}.

Thus the monopoles of type α and charge λ will have the same parameter space for all the hyperbolic spaces $H^3(p^{-1})$ with $p=1, 2, 3, \ldots$. As $p \to \infty$ the hyperbolic space $H^3(p^{-1})$ tends to the flat space R^3. This suggests that perhaps the same parameter space also applies to monopoles on R^3. This conjecture has now been proved by Donaldson [8] for $G=\mathrm{SU}(2)$, though his argument is direct and does not use the hyperbolic space.

The conjecture we have just put forward would imply that every monopole on R^3 is, in a natural way, a limit of k-instantons (invariant under S^1) as $k \to \infty$. The explicit formulae for instantons are all rational while monopole formulae tend to involve exponential functions. The limiting procedure involves therefore expressing an exponential as a limit of rational functions of increasing degree. Chakrabarti [5] has been deriving such formulae for monopoles as limits of instantons, and it would be interesting to relate his computational approach to our more theoretical one. It seems likely that both use the same basic mechanism.

A more detailed investigation of monopoles on hyperbolic space paralleling Hitchin's approach will be given in [2].

Finally it may be worth making some general comments on the relation between monopoles and rational maps. Already in Sect. 3 we mentioned the theorem of Segal [19] about the topology of spaces of rational maps, namely that in the limit when the degree tends to ∞ this space has the homotopy type of the relevant space of continuous maps ([19] deals with maps $CP_1 \to CP_n$, but the methods have been generalized to cover other cases than CP_n). It follows that the parameter space of monopoles has (conjecturely) similar properties. Now this is a result which might be susceptible to direct analytical proof by Morse theory. In fact Taubes [20] has already taken steps in this direction by showing that [for SU(2)] there are non-minimal solutions of the Yang-Mills-Higgs equations. By contrast it is known that there are no non-minimal harmonic maps $CP_1 \to CP_1$: Morse theory just fails to work here (the equations are of "critical-exponent" type). It seems therefore that the Yang-Mills-Higgs equations in R^3 can be viewed as some sort of "regularization" of the harmonic map equation on S^2. This could eventually provide an analytical explanation of Segal's theorem. Moreover, because of the analogy between instantons in 2D and 4D which has been the main point of this paper, it is tempting to speculate on possible equations in R^5 which might similarly "regularize" the Yang-Mills equations on S^4.

Acknowledgement. I am indebted to S. K. Donaldson and N. J. Hitchin for many helpful discussions. The idea that ΩG should be considered in connection with instantons I owe essentially to H. Garland.

References

1. Atiyah, M.F.: Geometry of Yang-Mills fields. Lezioni Fermiane, Scuola Normale Superiore, Pisa, 1979
2. Atiyah, M.F.: Magnetic monopoles on hyperbolic space. Proc. International Colloquium on vector bundles, Tata Institute, Bombay 1984 (to appear)
3. Atiyah, M.F., Drinfeld, V.G., Hitchin, N.J., Manin, Yu.I.: Construction of instantons. Phys. Lett. **65**A, 185–187 (1978)
4. Atiyah, M.F., Jones, J.D.J.: Topological aspects of Yang-Mills theory. Commun. Math. Phys. **61**, 97–118 (1978)
5. Chakrabarti, A.: Instanton chains with multimonopole limits: Lax pairs for non-axially-symmetric cases. Phys. Rev. D**28**, 989 (1983)
6. Donaldson, S.K.: Anti-self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles. Proc. Lond. Math. Soc. (to appear)
7. Donaldson, S.K.: Instantons and geometric invariant theory. Commun. Math. Phys. (to appear)

8. Donaldson, S.K.: Nahm's equations and the classification of monopoles (to appear)
9. Forgacs, P., Horvath, Z., Palla, L.: An exact fractionally charged self-dual solution, Hungarian Academy of Sciences. Preprint KFKI 60 (1980)
10. Hitchin, N.J.: Monopoles and geodesics. Commun. Math. Phys. **83**, 579–602 (1982)
11. Jaffe, A., Taubes, C.H.: Vortices and monopoles. Boston: Birkhäuser 1980
12. Kirillov, A.A.: Elements of the theory of representations. Berlin, Heidelberg, New York: Springer 1978
13. Maruyama, M.: Stable vector bundles on an algebraic surface. Nagoya Math. J. **58**, 25–68 (1975)
14. Nahm, W.: All self-dual multimonopoles for arbitrary gauge groups (preprint), TH. 3172 – CERN (1981)
15. Pressley, A.N.: The energy flow on the loop space of a compact Lie group. J. London Math. Soc. (to appear)
16. Pressley, A.N.: Decompositions of the space of loops on a Lie group. Topology **19**, 65–79 (1980)
17. Pressley, A.N., Segal, G.B.: Loop groups. Oxford: Oxford University Press 1984
18. Segal, G.B.: Unitary representations of some infinite-dimensional groups. Commun. Math. Phys. **80**, 301–342 (1981)
19. Segal, G.B.: The topology of spaces of rational functions. Acta Math. **143**, 39–72 (1979)
20. Taubes, C.H.: The existence of a non-minimal solution to the SU(2) Yang-Mills-Higgs equations in $\mathbb{R}^3$. Commun. Math. Phys. **86**, 257–298, 299–320 (1982)

Communicated by A. Jaffe

Received January 30, 1984